

**The Value of the Information That Can Be Generated: Optimizing Study Design to Enable
the Study of Treatments Addressing an Unmet Need for Rare Pathogens**

Aaron Dane¹, John H.Rex², Paul Newell², Nigel Stallard³

¹DaneStat Consulting, Macclesfield, UK, ²F2G Limited, Manchester, UK, ³Warwick Medical
School. University of Warwick, UK.

Corresponding author: Aaron Dane, Cheshire, UK. Email: aarondane@danestat.com, +44 7821
720 631

Alternate Corresponding Author: John Rex, Massachusetts, USA. Email: john.rex@f2g.com

Keywords: antibiotics; antifungals; unmet need; small trials; alternate statistical criteria.

Abstract

In traditional phase 3 trials confirming safety and efficacy of new treatments relative to a comparator, a one-sided type I error rate of 2.5% is traditionally used, and typically leads to minimum sizes of 300-600 subjects per study. However, for rare pathogens, it may be necessary to work with data from as few as 50–100 subjects. For areas with a high unmet need, there is a balance between traditional type I error and power and enabling feasible studies. In such cases, an alternative one-sided alpha level of 5% or 10% should be considered and we review herein the implications of such approaches. Resolving this question requires engagement of patients, the medical community, regulatory agencies, and trial sponsors.

Introduction

Traditional phase 3 trials confirm safety and efficacy of a new treatment relative to an active comparator or placebo using a one-sided type I error rate of 2.5%. Although the 2.5% type I error rate is not a regulatory requirement, it is traditionally used by regulatory agencies to limit false conclusions of treatment effect. Given the high unmet need, in some infectious disease research FDA have considered flexibility in benefit-risk considerations, granting approval with a single adequate and well-controlled trial, with supportive information such as a phase 2 trial or information from in vitro studies and animal models of infection. Although this enables progression of some clinical development programs, there are still situations when it may only be feasible to recruit 50-100 patients in many years (1), for example the plazomicin bacteremia study for carbapenem-resistant enterobacteriaceae (2), making even a single trial extremely challenging. Resultingly there are no, or very limited, data in such patient populations, so development of agents in small populations thus requires alternative approaches. Importantly, we only envision these approaches being used when a larger trial is simply not possible, so these ideas do not apply when there is a realistic choice between a larger or smaller study.

Drawing on value of information ideas within the orphan drug area (3-6), we propose an approach that uses alternative type I error rates in a decision-making and sample size determination framework where typically sized controlled studies are extremely challenging to enroll. The goal is to balance the risks of correct and incorrect decisions from limited data. Both approval and non-approval carry risks and we propose an approach which considers risks of failing to approve a new effective medicine, approval of an ineffective medicine, and conducting a large RCT (with some patients receiving sub-optimal therapy) in a limited population.

This paper discusses regulatory approval on the basis of efficacy, as the aim is to understand the risks of using alternative statistical criteria within an efficacy assessment. The safety and benefit-risk profile are also critically important and need to be considered in any evaluation, but as the safety profile is unique to

each treatment, we have not discussed this aspect in detail, but it is always a critical aspect of any evaluation of the evidence to support regulatory approval.

Superiority trials are not appropriate for all development programs

It is often perceived that demonstration of superiority in small trials with resistant pathogens should be straightforward if a strong effect exists. However, this is infeasible because we cannot study ineffective comparators in seriously ill patients, and demonstration of superiority is challenging when therapies exist with some efficacy, albeit suboptimal (7, 8). Although combinations of less effective therapies may have toxicities, their existence makes superiority hard to demonstrate.

In some settings, the use of a non-standard significance level to assess superiority may make studies more feasible, but in many cases the issues associated with demonstrating superiority described above will still exist, meaning alternative options are also required.

In this paper we therefore focus on non-inferiority trials, but similar arguments to those presented could be used to assess the impact of changes to type I error in superiority trials.

Decision-making and sample size in RCTs with rare pathogens

The challenge reduces to recognizing that rare pathogens make it impossible for practical purposes to conduct either a small superiority study or a large non-inferiority active-control study. And yet, new therapies are needed for rare pathogens, and we can approach this problem by being sure that we are confident of the following:

- We will approve effective treatments (showing superiority or non-inferiority with an active control).
- We will not approve ineffective treatments.

1 In registrational studies it is traditional to set power (probability of correctly showing the new agent is
2 effective) at 80-90% and type I error at 2.5% (one-sided, the probability of concluding a new treatment is
3 effective when it is not)) and set sample size accordingly (9). Table 1 summarizes these concepts and the
4 consequences for patients within the trial and those beyond the trial who will receive the current or new
5 therapy dependent upon the trial results.

6 When we design the trial, we do not know the true response rate, so must understand the probability of
7 approval for different study sizes and scenarios regarding true treatment effects. We are attempting to
8 identify a “sweet spot” for sample size balancing the type I error and power when it is infeasible to obtain
9 the traditional chances of correctly or incorrectly approving a new treatment.

10 Where possible, trials with conventional power and type I error are preferable. However, if the trial is too
11 large this deprives patients of timely access to an effective therapy and means that whilst the trial is
12 ongoing some patients within the trial, will receive an inferior treatment. If a trial is too large developing
13 the therapy is infeasible as we cannot generate data for that patient population. Conversely, too small a
14 trial will give a larger probability of making an incorrect decision (either approval or non-approval).

15 We propose that when large trials are infeasible, alternative type I error rates should be explicitly
16 considered within the context of the benefits and risks of new treatments, particularly when accompanied
17 by supportive information such as a data from in vitro studies and animal models of infection. We
18 consider this for rare pathogens where new therapies are needed, and patient numbers make clinical trials
19 prohibitively lengthy or simply impossible to undertake. This has the potential to speed availability of
20 new effective treatments rather than condemn patients to suboptimal treatment while larger trials are
21 either attempted in vain or never attempted.

Understanding Clinical Trial risks

In considering alternative statistical criteria, we first state key risks and then consider acceptable risk levels in an area of extreme unmet need. Risks must be considered for a range of scenarios where the true effect of the test agent related to control is unknown. This is true of all studies but may need to be considered differently when trials cannot be designed to standard levels of power and type I error. Important scenarios are situations when the new agent is better than, worse than, or similar to control. We consider these in turn for a randomized clinical trial using a 1:1 randomization to new agent or control with a 20% NI margin and 95% CI (two-sided $\alpha=0.05$), as has been used for some trials (10, 11).

Scenario #1: New agent better than control

Imagine the new agent and control have true response rates (unknown before the trial is conducted) of 60% and 40% respectively. The correct outcome would be to conclude noninferiority (NI) and approve the new agent. The probability that NI is concluded (power) with this scenario increases with sample size (N) (Figure 1A) but does not improve greatly with more than 40-50 patients/group, when power approaches 90%. Although more data may be needed from a safety perspective this would be sufficient to show the new agent has efficacy NI to control.

Scenario #2: New agent slightly better than control

When the new agent and control have true response rates of 50% and 40% respectively the correct outcome would again be to conclude noninferiority (NI) and approve the new agent. In this case, as presented in Figure 1B, the probability NI is concluded increases with sample size, but power is lower for a given sample size than for Scenario #1 and approaches 80% at 50 patients/group.

Scenario #3: New agent similar to control

In the third scenario the true (but as yet unknown) response rates are the same (40%) for new agent and Control. The correct decision in this scenario is more nuanced and must consider the value of currently

1 available therapies, even if these are suboptimal. As stated, we assume that it would be desirable to have
2 additional therapies available.

3 In this case the probability of concluding NI (power) increases with sample size (Figure 1C) but requires
4 a larger sample size. A study with ~50 patients/group has ~50% power, whilst a study of 80
5 patients/group provides ~80% power. It would take a large study to achieve 90% power. As it is unlikely
6 that studies would be undertaken with only a 50/50 chance of success, developing new agents is not
7 possible when it is infeasible to recruit more than 100 patients. If new therapies are badly needed, a
8 discussion on required levels of type I error and power should be considered.

9 Scenario #4: New agent slightly worse than control

10 When the true response rates for the new agent and control are 30% and 40% respectively, the correct
11 outcome and associated approval is less obvious. If the primary endpoint of the study relates to all-cause
12 mortality, you would not want to approve such an agent, but with a response rate endpoint, if there is a
13 high unmet need, the available treatment has significant toxicity and there is increasing resistance to this
14 treatment there could still be value in a new agent. You would not always want to approve such an agent,
15 but consideration of these components of the available therapy would be important when considering the
16 need for new agents. Therefore, with this scenario there is no correct outcome for every situation. Rather,
17 consideration of whether the new agent should or should not be approved will depend upon the reason
18 why a new agent is needed. In this case the plot shows the probability of concluding NI is < 25% for
19 sizes of < ~75/arm (Figure 1D). The probability of concluding NI increases as the sample size further
20 increases and at 200/arm, the probability approaches 50%. However, such sample sizes are presumed
21 impossible in this setting. Net, this again shows a relatively small chance of concluding NI with
22 achievable sample sizes of 40-50 patients/group.

Scenario #5: New agent worse than control

When the true (but as yet unknown) response rates for the new agent and control are 20% and 40% respectively, we assume we would definitely not want to approve the new agent, so the correct outcome is that the new agent should not be approved. In this case the probability of concluding NI is always very low (Figure 1E). As the true difference between treatments equals the NI margin this is the Type I error rate (2.5%), so we rarely conclude NI erroneously, whatever the study size. Therefore, 40-50 patients/group again appears reasonable.

The effect on future patients with this rare condition

We have shown that trials studying infections caused by rare pathogens are infeasible with traditional power and Type I error rates. Another important component is the implication of any drug approval decision on the treatments received by patients with this rare pathogen, both in the RCT and in the future. As a hypothetical example, suppose we have an overall population of 1,000 patients who might ever benefit from a new treatment, and from this population we recruit 100 patients (50/arm) to a trial comparing the new agent with Control, the results of which will define the therapy used for the remaining 900 patients, dependent upon the NI conclusion made from the trial (see Table 2).

Assuming all patients are included in the trial or are treated based on the results of the trial, we can calculate how many patients are successfully treated using the following:

- How many of 50 trial patients on the new agent from trial are successfully treated?
- How many of 50 trial patients on Control are successfully treated?
- How many of remaining 900 patients are successfully treated depending upon the trial outcome?

The expected number of successes can be calculated under the scenarios considered above. When the new agent response rate is 60%, we would expect 30 of 50 trial patients receiving the new agent and 20 of 50 trial patients receiving control to respond (Table 2). We would expect 540 of the remaining patients to

respond if NI is concluded and the new agent is used, and 360 of the 900 patients to respond on Control if NI were not concluded.

Using a 20% NI margin and 95% CI, with true response rates of 60% and 40% a trial with 50 patients/arm would have a 98.3% chance of concluding NI. Thus, we expect the new agent to be approved 98.3% of the time (so 540/900 future patients would respond) and the Control to be used (so 360/900 respond) 1.7% of the time. The total expected number of responses from the 1,000 patients is therefore 587 as displayed in the first row of *Table 2*. The second row of Table 2 shows similar calculations using the same assumptions, but a larger RCT including 200 patients/group, leaving 600 patients treated based on the RCT results. In this case, the total expected number of responses is lower at 560 patients. This may seem counter-intuitive but although larger RCTs increase the probability of reaching the correct conclusion, they also mean we wait longer to reach this conclusion meaning more patients within the trial are randomized to receive a less effective therapy, and a smaller number of patients benefit from the more effective therapy after the trial.

The example in Table 2 is for the most optimistic scenario considered and presents results for trials with 50 and 200 patients per arm only. Similar calculations are possible for each of the five scenarios described and for different sample sizes and show a similar pattern of declining response with larger sample sizes for less optimistic scenarios (Figure 2). Recall that Figure 1 shows the probability of concluding NI in each case, whilst Figure 2 shows the proportion of responses expected in the population for each scenario. These results can be used to assess whether the risks of correctly approving an effective therapy, incorrectly approving an ineffective therapy, and the number of patients receiving suboptimal treatment, have been appropriately balanced for a given trial size.

If the new agent is better than the control, a trial with 40 patients/group gives reasonable power and optimizes the expected successes in the overall population based on the treatment patients receive during and after the trial (Figure 2A). When the new agent is slightly better than control the power is lower than when the new agent is clearly better, but still declines after 40-50 patients/arm.

1 If the new agent is worse than Control, as noted above, the probability of incorrectly recommending the
 2 new agent is low irrespective of the RCT sample size (Figure 2E). The expected proportion of responses
 3 decreases with increasing sample size as more patients in the trial receive the (less effective) new agent.
 4 When the new agent is slightly worse than control, a similar pattern is seen with the expected proportion
 5 of successfully treated patients, whereas the probability of concluding NI increases with increasing
 6 sample size, but there is still only 15-20% probability of concluding NI with 40-50 patients/arm.

7 If the new agent and Control have similar efficacy, the probability of recommending the new agent
 8 increases with N, and a larger study is required using traditional criteria (Figure 2C). In this scenario,
 9 because the new agent and Control have the same response rate the expected proportion of patients to
 10 respond is constant (here 40%). Although the unmet need may appear less urgent when the efficacy of the
 11 new agent is similar to control, there are still a number of cases where additional treatment options are
 12 urgently needed due to emerging resistance or toxicity concerns with current therapies. In this scenario a
 13 study including ~100 patients/group would be required to demonstrate efficacy. In studies of some rare
 14 or resistant pathogens even 40-50 patients/group may be challenging, so it is necessary to find a way of
 15 running a feasible trial. We next suggest a framework for assessing alternative statistical criteria building
 16 on ideas presented above.

17 **A framework for using alternative statistical criteria**

18 We have shown that although conventional error rates are achieved using 40-50 patients/ arm when the
 19 new treatment is truly clearly better or worse than the control, 100 patients/arm may be necessary when
 20 the two treatments have similar efficacy. When such a study is infeasible, and additional treatment
 21 options are desirable, use of an alternative type I error rate (or confidence interval) can be considered.
 22 Here we present the use of an 80% confidence interval (two-sided $\alpha=0.2$) and a 20% NI margin.

1 This approach would be reserved for circumstances with high unmet need and limited population where it
2 is critical that new therapies are available but could not otherwise be developed. We will not discuss how
3 an agent would meet these criteria but do outline how the risks associated with failure to make new
4 therapies available, and the risk of incorrectly approving new agents could be considered for such agents.

5 Properties of trials with an 80% CI are shown in Figure 3, which includes in one plot the type of
6 information presented in Figures 1 and 2. When the new agent is better than control, the expected number
7 of patients responding declines after ~50 patients/arm. The probability of a successful trial when the
8 treatments are similar is higher (panel C), approaching 80% with 50-60 patients/arm. The probability of
9 such a trial incorrectly concluding NI when the new agent is worse (panel E) is 10%, so above the usual
10 value of 2.5% when using a 95% CI. This should be a point for discussion between patients, ID
11 physicians, regulatory agencies, and sponsors for rare pathogens or for limited populations when
12 additional treatment options are desperately needed. When the new agent is slightly worse a similar
13 pattern is seen to that when using a 95% CI in that the probability of concluding NI increases with
14 increasing sample size, in this case meaning the chance of concluding NI approaches 40% with 50
15 patients/arm. As stated previously, the utility of an agent with the possibility of a slightly lower response
16 rate, but the ability to provide other benefits in terms of less toxicity or pathogen resistance should form
17 part of the assessment of using an 80% CI.

18 The previous sections showed that when the new agent is clearly superior to the control this would be
19 apparent from 50 patients/arm. The alternative type I error rate level is important is when the new agent
20 and control are similar or inferior. If it is only feasible to run a study of 100 patients, a trial with
21 traditional type I error rate is impossible. Additional flexibility is then required, particularly when
22 available therapies are sub-optimal, such as being the drug of last resort or with toxicity issues where it is
23 desirable to have additional therapies available. In this case it will often be a choice between a smaller
24 dataset or having no data at all.

Discussion

Although studies of rare pathogens are challenging, the development of agents for such pathogens remains important. We have shown that while studies with traditional statistical criteria may be impossible, we can articulate a decision-making framework to support smaller studies. The proposed framework displays the risks of incorrect approval of an ineffective agent and incorrect rejection of an effective agent based on efficacy, along with the expected number of successes within a small population, highlighting that agreement needs to be reached across the community on how to balance these factors. The approach is similar to that proposed for orphan drug trials (3-6) where a type I error rate larger than 0.025 is sometimes used (12, 13).

We have shown that comparative trials with 50 patients/arm can provide reliable decisions and could be the key to enabling any development for patients with rare pathogens. In our example we have assumed that the control group event rate is 40%, and although our exact findings regarding the probability of approval may not hold if the control group rate were very different, similar methods could be used to choose an appropriate sample size.

We have also shown that a larger study does not always provide better outcomes in a limited population because fewer patients remain to benefit from a therapy following the trial. Also, in a trial with a 1:1 randomization where one treatment is less effective, 50% of patients receive this sub-optimal therapy.

We have presented an example where an 80% confidence interval is used alongside a 20% NI margin. Dependent upon the degree of unmet need, feasibility of recruitment and the basis for justifying the NI margin, alternative approaches using different NI margins and levels of alpha (such as 90% CIs) could be considered to make trials more feasible. We acknowledge that the use of 80% or 90% confidence intervals increases the risk of falsely concluding non-inferiority when compared with the traditional 95% CI, and this is the reason this approach should be reserved for situations where new treatments are urgently needed for a small patient population where a trial using a 95% CI is not possible. In these

1 situations, additional flexibility may be warranted and would maintain scientific credibility by explicitly
2 stating the error risk levels and would be particularly justified in the situation we envisage where the trial
3 results are also augmented with robust information from in vivo and in vitro models of infection. We feel
4 that this is also aligned with the intention stated in FDA guidance, where FDA can approve if there is “a
5 *positive benefit-risk balance in the limited population, ... even though insufficient data exist to conclude*
6 *that there is a favorable benefit-risk profile in a broader population.*”

7 Our calculation of patients benefitting from the chosen therapy assumed that all patients are either
8 included in the trial or benefit from its results by receiving the preferred treatment. In reality, some
9 patients will receive standard treatment outside the trial while it is ongoing. Including these patients in
10 the calculations of the expected proportion of patients responding would lead to even smaller trials being
11 optimal when the new agent was effective (as even more patients are receiving a sub-optimal therapy) and
12 would make little difference when it was not effective. Therefore, as the trial sample size becomes large
13 the expected number of patients benefitting from the chosen therapy would decrease even more rapidly
14 than in the calculations presented in this paper.

15 It is acknowledged that this is a different level of evidence of efficacy than is usually used for regulatory
16 approval, and this is why it is critical to gain the agreement of regulatory agencies and the scientific
17 community when designing such studies that there is sufficient unmet and a small population which
18 justifies this alternate approach to provide interpretable data for this patient population what would not
19 otherwise be available. Further, and as with any NI trial, it is important to be sure of the effectiveness of
20 the control agent, and that the primary endpoint is objective and reflects a true clinical benefit to patients.

21 Beyond a change in statistical criteria for the primary endpoint, agreement to such a design and any final
22 assessment would always consider the benefit-risk profile of a new agent. This would include detailed
23 review of secondary efficacy endpoints. Knowledge of safety risks of an agent before the trial, or with a
24 specific drug class, may impact the acceptability of a smaller safety dataset. These considerations are

specific to a given agent but would always need to be considered when designing a trial as we have proposed.

This approach is relevant both when a drug has activity only against a specific pathogen(s) or a drug with broad-spectrum activity that includes an uncommon pathogen. In both cases this would importantly provide quantitative data on pathogens of interest, with the key difference being the amount of additional information it is possible to generate.

In conclusion, registrational trials for rare pathogens are difficult, and sometimes impossible, to conduct using traditional statistical criteria. Explicit consideration by the medical, regulatory, and patient community of the ability to correctly approve an effective therapy, the risk of incorrectly approving an ineffective therapy, and the overall number of patients expected to benefit should be used to help define clear success criteria to ensure such studies can be conducted, so data are available in these important patient groups.

Potential Conflicts of Interest

Aaron Dane acts as a paid consultant for the Pharmaceutical and Biotechnology Industry. Companies include Achaogen, Allegra, Amicrobe, Amplyx, AN2, Artizan, Cidara, ContraFect, Correvio, Davolterra, Destiny, F2G, Entasis, Geom, GSK, Gyroscope, Humanigen, Kymab, Mironid, Modis, Nabriva, Orca, Pfizer, Phico, Pled, Polyphor, Roche, Scynexis, Sinovent, SNIPR Biome, Spero, TenNor, Transcrip, VenatoRx and Zavante.

During the period 2018-2021, John H. Rex, MD is/has been Chief Medical Officer & Director, F2G, Ltd., Editor-in-Chief, AMR.Solutions, Operating Partner & Consultant, Advent Life Sciences, and Adjunct Professor of Medicine, McGovern Medical School, Houston, TX; has received grant support from Wellcome Trust; sits (or sat) on the scientific advisory boards of Bugworks Research, Inc., Basilea Pharmaceutica, Forge Therapeutics, Inc., Novo Holdings, Roche Pharma Research & Early Development,

1 and Sumitovant; received consulting fees from Forge Therapeutics, Inc., Innocoll, Vedanta, Progenity,
2 Nosopharm SA, Roivant Sciences, Shionogi Inc., GlaxoSmithKline, and Pfizer Pharmaceuticals. He is
3 currently a shareholder in AstraZeneca Pharmaceuticals, F2G, Ltd, Advent Life Sciences, Zikani
4 Therapeutics, and Bugworks Research, Inc. The opinions expressed are his own and do not necessarily
5 reflect the opinion of any of the groups with which he works.

6 Paul Newell is a paid employee of F2G Ltd, has share options in F2G Ltd and is a shareholder in
7 AstraZeneca Pharmaceuticals.

8 Nigel Stallard has no potential conflicts of interest to declare.

9 **Patient consent**

10 This study did not require obtaining of patient consent.

References

1. Rex JH. FDA AMDAC – Developing single-pathogen agents For *P. aeruginosa* and *A. baumannii*. 2017; Accessed online on 31 Oct 2020 at <https://amr.solutions/2017/04/13/fda-amdac-developing-single-pathogen-agents-for-p-aeruginosa-and-a-baumannii/>.
2. McKinnell JA, Dwyer JP, Talbot GH, Connolly LE, Friedland I, Smith A, et al. Plazomicin for Infections Caused by Carbapenem-Resistant Enterobacteriaceae. *New England Journal of Medicine*. 2019;380(8):791-3.
3. Willan AR, Pinto EM. The value of information and optimal clinical trial design. *Stat Med*. 2005;24(12):1791-806.
4. Abrahamyan L, Willan AR, Beyene J, McLimont M, Blanchette V, Feldman BM, et al. Using value-of-information methods when the disease is rare and the treatment is expensive--the example of hemophilia A. *J Gen Intern Med*. 2014;29 Suppl 3:S767-73.
5. Pearce M, Hee SW, Madan J, Posch M, Day S, Miller F, et al. Value of information methods to design a clinical trial in a small population to optimise a health economic utility function. *BMC medical research methodology*. 2018;18(1):20.
6. Miller F, Zohar S, Stallard N, Madan J, Posch M, Hee SW, et al. Approaches to sample size calculation for clinical trials in rare diseases. *Pharm Stat*. 2018;17(3):214-30.
7. Rex JH, Talbot GH, Goldberger MJ, Eisenstein BI, Echols RM, Tomayko JF, et al. Progress in the fight against multidrug-resistant bacteria 2005-2016: Modern non-inferiority trial designs enable antibiotic development in advance of epidemic bacterial resistance. *Clinical Infectious Diseases*. 2017;65:141-6.
8. Rex JH, Eisenstein BI, Alder J, Goldberger M, Meyer R, Dane A, et al. A comprehensive regulatory framework to address the unmet need for new antibacterial treatments. *Lancet Infect Dis*. 2013;13(3):269-75.
9. Chow S-C, Wang H, Shao J. *Sample Size Calculations in Clinical Research* (2nd ed.): Chapman and Hall/CRC: <https://doi.org/10.1201/9781584889830>; 2007.
10. Entasis Therapeutics Holdings Inc. Form 10-K. Accessed 4 May 2022 at https://www.annualreports.com/HostedData/AnnualReports/PDF/NASDAQ_ETTX_2020pdf. 2020.
11. Cidara Therapeutics. Press Release: Mundipharma and Cidara Therapeutics announce first presentation of results from global Phase 3 restore trial of rezafungin for treatment of candidemia and/or invasive candidiasis demonstrating its positive efficacy and safety profile. Accessed 4 May 2022 at <https://ircidaracom/news-releases/news-release-details/mundipharma-and-cidara-therapeutics-announce-first-presentation>. 2022.
12. Lagakos SW. Clinical trials and rare diseases. *N Engl J Med*. 2003;348(24):2455-6.
13. Gallin JI, Alling DW, Malech HL, Wesley R, Koziol D, Marciano B, et al. Itraconazole to prevent fungal infections in chronic granulomatous disease. *N Engl J Med*. 2003;348(24):2416-22.

1 *Table 1: Scenarios to consider when analyzing decision-making with small datasets.*

True response rate or mortality rate	Consequences within RCT	Consequences for patients after RCT
New agent is worse than Control	Every patient randomized to New Agent risks a worse outcome	If the new agent is approved, this problem is perpetuated, and all future patients would receive a less effective therapy <i>We should avoid incorrect approvals so want to have low power of approval</i>
New agent is better than Control	Every patient randomized to Control risks a worse outcome	If the new agent is <u>not</u> approved, this problem is perpetuated, and all future patients would receive a less effective therapy <i>We should retain high power of approval</i>
New agent is similar to Control	Every patient has approximately the same outcome	Either existing treatment or a new therapy would be acceptable, but we still need to have additional therapies available. <i>Therefore, we still want to retain high power of approval</i>

2

3

- 1 *Table 2: Expected Number of Responses in the Entire Population with an NI trial of 100 or 400 patients:*
- 2 *Response Rates of 60% for the new agent v 40% for Control*

RCT Sample Size	Responses in RCT		Probability New Agent is approved	Responses Following Trial		
	New Agent	Control		If New Agent is approved		If New Agent not approved (Control is used)
50/arm	30/50	20/50	0.983	540/900		360/900
<i>Expected successes in overall population:</i> $30 + 20 + (0.983 \times 540 + (0.017 \times 360)) = 587$ (58.7% of all patients)						
200/arm	120/200	80/200	0.999	360/600		240/600
<i>Expected successes in overall population:</i> $120 + 80 + (0.999 \times 360 + (0.001 \times 240)) = 560$ (i.e. 56% of all patients)						

Figure 1: Plot of power when using 95% CI and -20% NI margin

Figure 2: Plot of expected proportion of responses in a trial using 95% CI and -20% NI margin

Figure 3: Plot of power and expected proportion of responses in a trial using 80% CI and -20% NI margin

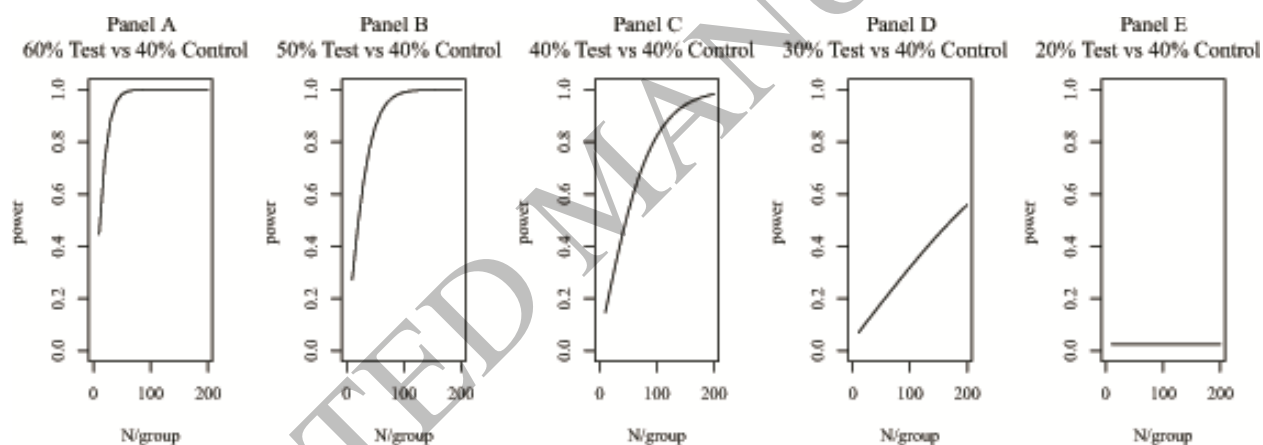


Figure 1
200x69 mm (.23 x DPI)

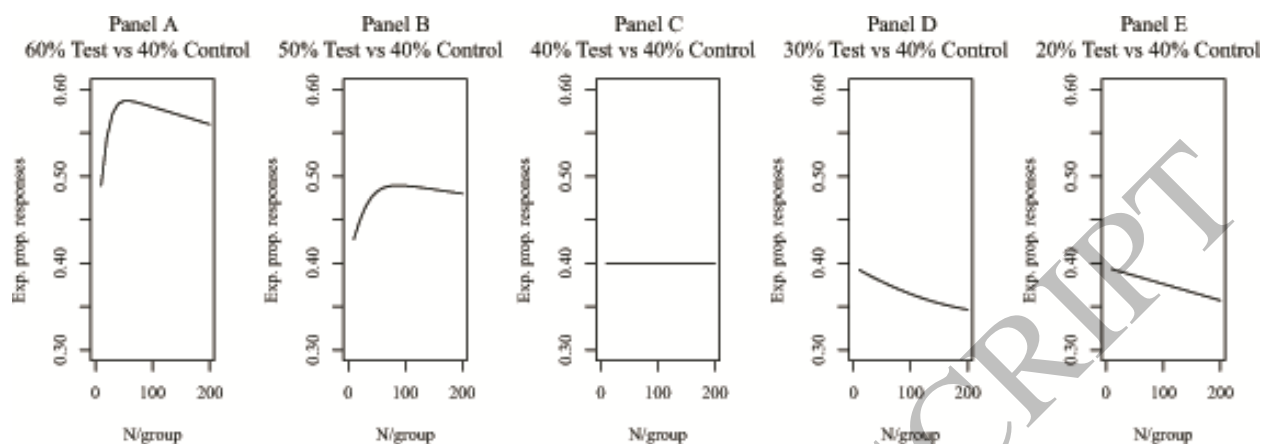


Figure 2
201x69 mm (.23 x DPI)

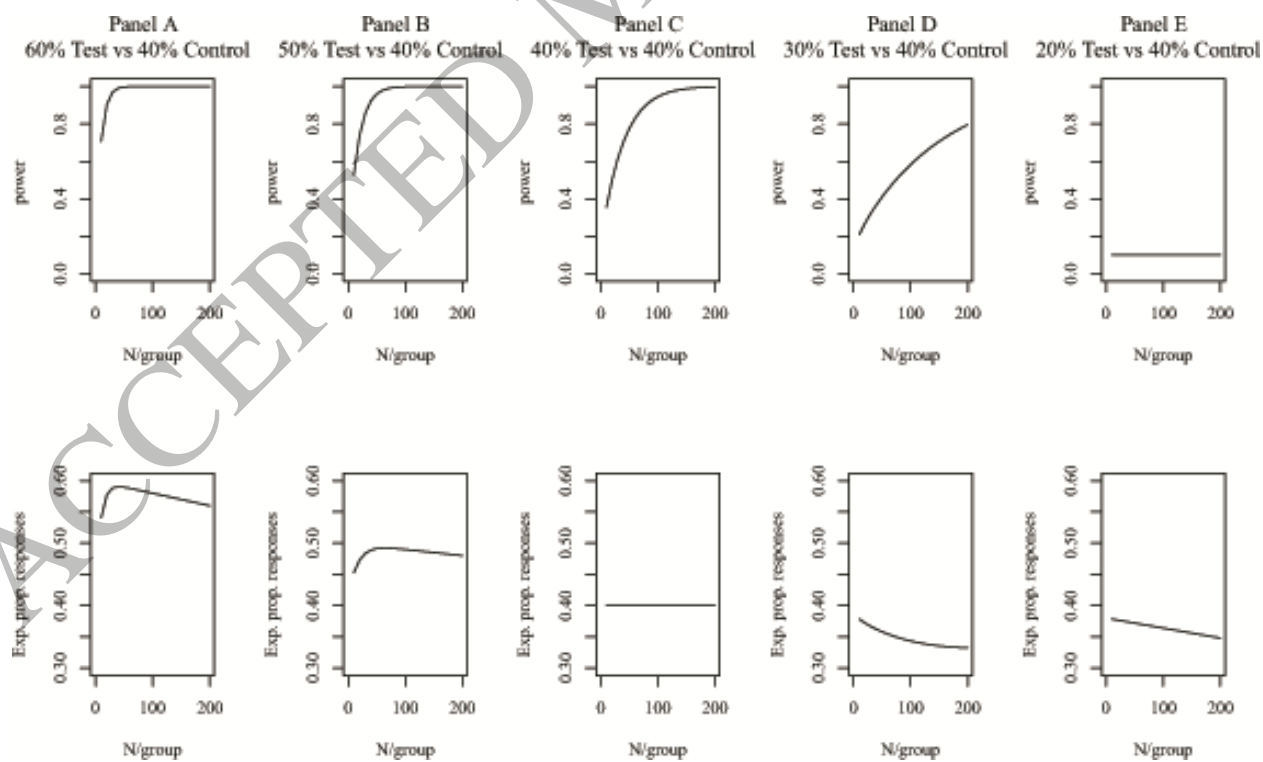


Figure 3
201x120 mm (.23 x DPI)